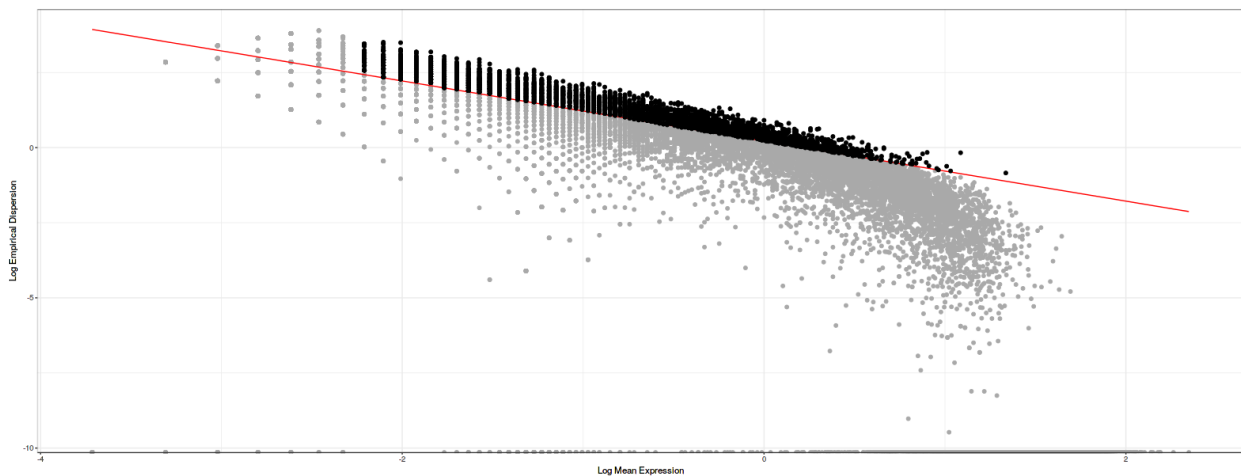


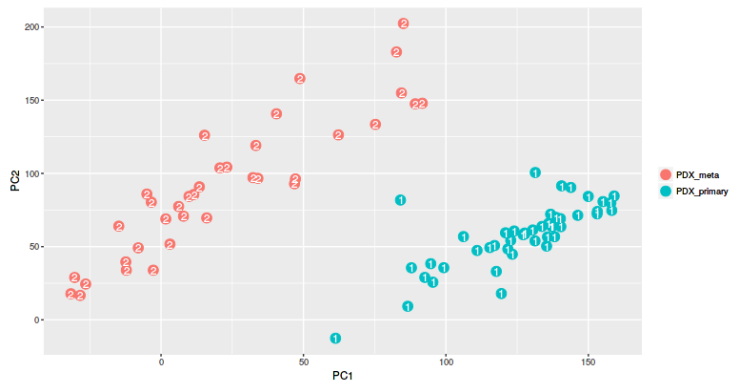
Suppl. Figure 1: Granatum total running time(in log10 scale)with various numbers of cells.

Datasets with various sizes from two single-cell platforms (Fluidigm C1 and 10x Genomics) are used. To generate expression data up to 6000 cells, the Fluidigm C1 datasets are simulated using Splatter, with parameters estimated from the K-dataset (118 cells). The 10x Genomics datasets are down-sampled from the original 6000-cell PBMC dataset. The x-axis represents the size of the dataset, and the y-axis represents the total running time (in minutes) of Granatum. The imputation step (scImpute) takes about 90% of the total running time, and Monocle based pseudo-time construction step takes about 80% of the rest of the running time.

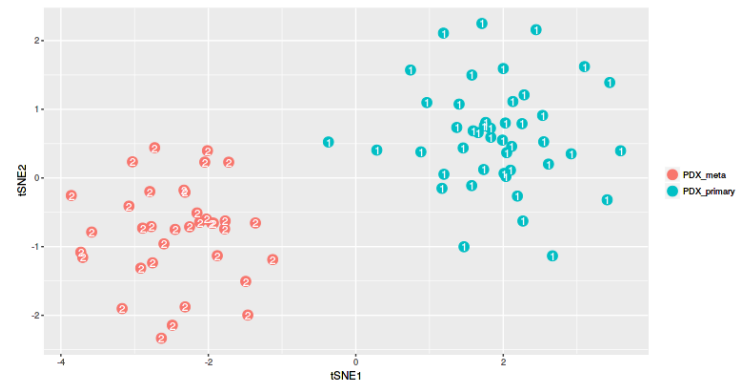


Suppl. Figure 2: The Gene filtering step. The y-axis of the scatter-plot is the empirical dispersion, estimated by a negative binomial model. The x-axis is the log mean expression of each gene. The red line is the fit of a negative binomial model onto the data. Black points represent gene to be kept and gray points are filtered genes.

A



B



Suppl. Figure 3: The Clustering step. (A) PCA and (B) Correlation t-SNE plots of single cells (dots) are shown, with colors indicating the cell types reported in the original dataset and cluster number (1,2) super-imposed on the cells.